Aula 7

# AMOSTRAGEM: POPULAÇÃO E AMOSTRA. TIPOS DE AMOSTRAGEM. AMOSTRA PILOTO. NÍVEL DE CONFIANÇA. ESTIMATIVA DA MÉDIA E PROPORÇÃO POPULACIONAL POR PONTO E POR INTERVALO.

## META

Estudar características de populações com base nas informações colhidas por amostras de dados selecionados aleatoriamente nestas populações.

## OBJETIVOS

Ao final desta aula, o estudante deverá:

Saber o significado de população e amostra, bem como a diferença entre pesquisas censitárias e amostrais.

Entender o significado de nível de confiança e estimar parâmetros médios populacionais por ponto e por intervalo.

## PRÉ-REQUISITO:

Conhecimentos sobre estatística descritiva e probabilidades, principalmente sobre a distribuição normal. Também são importantes: Papel, Calculadora ou Computador para realização dos cálculos.

# INTRODUÇÃO

Olá! Tudo bem? Vamos dar seqüência ao estudo da estatística agora com a investigação sobre amostragem. Para isto é fundamental revisar todos os conceitos da estatística descritiva, principalmente os conceitos que envolvem cálculos sobre média aritmética e desvio padrão. Também são importantes os conhecimentos sobre probabilidade, principalmente aqueles relativos ao uso e aplicação da distribuição normal.

Quando optamos por pesquisas censitárias fazemos à escolha de pesquisar todos os elementos de uma população, ao passo que na investigação por amostragem pesquisamos apenas parte dos elementos aleatoriamente selecionados para comporem a amostra.

Embora em uma primeira impressão pareça difícil ou mesmo pouco preciso os resultados a ser obtidos por uma amostra, com o desenvolvimento desta aula você vai perceber que estes receios, caso eles existam, não têm fundamentos desde que na investigação de populações por amostragem sejam utilizados de forma precisa todos os critérios que envolvem o uso desta metodologia. Critérios estes que estão relacionados com todas as etapas de aplicação de uma amostra, desde o dimensionamento científico da amostra, passando pela seleção aleatória e investigação dos elementos sorteados até a análise dos resultados. Análise esta que tem por base as estimativas das cacterísticas populacionais fundamentadas no comportamento destas características na amostra desta população.

Geralmente o que parece ser mais simples em qualquer pesquisa seja ela censitária ou por amostragem é o que na maioria das vezes distorce alguns resultados. Deste modo sempre que venha a fazer qualquer pesquisa, em nosso caso por amostragem, tenha muito cuidado em determinar o tamanho ideal da amostra, na elaboração do questionário, na coleta dos dados, esta geralmente é a fase onde se concentra o maior volume de erros e por fim na crítica e análise dos resultados. Tendo estes cuidados podemos investigar características de populações por intermédio dos dados de uma amostra com uma precisão tão boa quanto a que seria obtida por uma pesquisa censitária

# AMOSTRAGEM ESTATÍSTICA

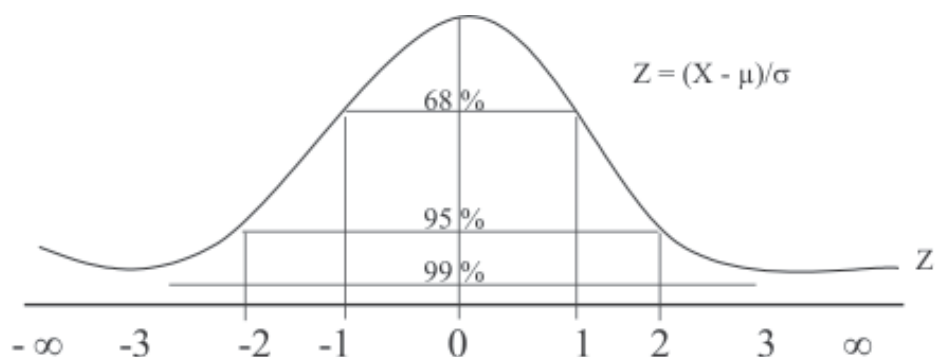
## NOÇÕES GERAIS

**Censo** – é a coleta de dados de todos os elementos que constituem um Universo ou População.
**Universo ou População** – é a totalidade dos elementos de um grupo, acerca dos quais se deseja fazer um determinado estudo. Os elementos da população devem possuir, pelo menos, uma característica comum. Exemplo: alunos da universidade, membros de sindicatos, servidores públicos, estabelecimentos (comerciais, industriais, agrícolas, etc,), atestados de óbitos, nascimentos, etc,.
Quanto ao número de elementos a população pode ser finita e infinita. Exemplo: população da cidade de Aracaju (finita), estudantes da UFS (finita), número de microorganismos em um lago (infinita), etc.
**Amostragem** – é o método de investigação de uma população por intermédio de uma pequena parte da mesma denominada amostra. Os principais tipos de amostragem são; amostragem probabilística e amostragem por quotas.
**Amostra** – é a fração representativa de uma população, devendo possuir todas as características da mesma tais como: média ($\mu$), proporção de ocorrência (p), variância ($\sigma^2$), etc. A amostra só deve diferir da população quanto ao número de elementos, visto que seus parâmetros devem ser bons estimadores dos parâmetros populacionais, para que o mesmo tenha representatividade.
**Amostragem Probabilística** – é uma amostragem científica, com fundamentos na curva normal e consequentemente no cálculo de probabilidades. Para uma população de tamanho N a probabilidade de participação de cada elemento na amostra é 1/N. Esta amostragem pode ser: Equiprovável, Estratificada e Por Conglomerados.
**Amostragem Equiprovavel ou Acidentalmente Simples –** quando os componentes da amostra são selecionados diretamente de um universo estudado, independente de seu tamanho ou de seu grau de variabilidade. Nesta amostragem todos os elementos do universo possuem a mesma probabilidade de participar da amostra. É um dos métodos mais utilizados, por ser bastante prático e eficiente, principalmente quando investigamos populações de características homogêneas.

Existem algumas restrições à sua aplicação, no caso de universos com grande variabilidade ou contido em vasta área territorial, principalmente relativo aos custos da pesquisa.

## NÍVEL DE CONFIANÇA

É uma medida estatística que traduz a probabilidade de ocorrência que o pesquisador espera obter nos resultados da pesquisa. A interpretação de um nível de confiança é feita em relação a curva normal. Quanto maior o nível de confiança adotado maior a probabilidade de acerto nas estimativas dos parâmetros populacionais.

De acordo com os estudos da curva normal padronizada, numa distribuição de médias amostrais de uma população; 68% das médias situam-se entre $\pm 1\sigma$, a contar da verdadeira média populacional, 95% situam-se entre $\pm 2\sigma$ e 99% entre $\pm 3\sigma$, aproximadamente.

Principais níveis de confiança utilizados em amostragem:

| Níveis de Confiança (%) | Coeficiente de Probabilidade Padronizado (z) |
|---|---|
| 90 | 1,65 |
| 95 | 1,96 |
| 95,5 | 2,00 |
| 98 | 2,33 |
| 99 | 2,58 |

## INTERVALO DE CONFIANÇA

Em amostragem probabilística os parâmetros média ( $\bar{x}$ ), proporção de ocorrência (p) e desvio padrão (s), quando resultantes de observação numérica, através de uma amostra, são estimados para a população por ponto ou por intervalo.

Na estimativa por ponto, não dispomos do grau de precisão destes parâmetros como estimadores dos parâmetros populacionais, visto que os valores observados na amostra são usados para estimar os dados populacionais sem nenhuma margem de segurança probabilística. ($\mu$ = x ; p = 'p e $\sigma$ = s).

Na estimativa por intervalo os parâmetros populacionais ($\mu$, p e $\sigma$) flutuam dentro de um intervalo de confiança com base no erro padrão ($s_x$ e $s_p$) dos respectivos parâmetros amostrais e no coeficiente de probabilidade (z) corresponde a um nível de confiança adotado, que por lógica deve ser o mesmo utilizado na amostra.

Dentro deste intervalo, supõe-se estar contido, com tantos por cento de probabilidade os parâmetros do universo estudado. Ou seja, qualquer parâmetro da amostra ( x, 'p e s ) deverá ter a mesma probabilidade de ser igual ao do universo de onde foi retirada a amostra.

A variância da população ($\sigma^2$) é encontrada a partir da fórmula:

$$\sigma^2 = \frac{\sum(X_i - \mu)^2}{n}$$

Quando a variância populacional é desconhecida pode ser estimada pela variância amostral ($\sigma^2$), sendo:

$$s^2 = \frac{\sum(x_i - x)^2 .}{n - 1}$$

Deve ter-se em conta que $s^2$ não é exatamente a estimação de $\sigma^2$, isto é; E $[s^2] \cong \sigma^2$. Esta estimativa só tem validade quando N é muito grande, fazendo com que N/(N–1) seja praticamente igual a 1. Caso contrario $s^2$ passa a ser uma estimativa viesada de $\sigma^2$. Neste caso para termos um estimador imparcial da variância populacional, com base na variância amostral, devemos seguir os seguintes critérios:

$$\sigma^2 = \frac{n}{n - 1} s^2$$ (quando a população for infinita)

$$\sigma^2 = \frac{n}{n - 1} \cdot \frac{N - 1}{N} s^2$$ (quando a população for finita).

Intervalo para estimativa da média ($\mu$) e da proporção de ocorrência (p) de um universo estudado, a partir de amostras com mais de 30 unidades.

Média ($\mu$)

$$P_r \ \{ \bar{x} - z\,\sigma_{\bar{x}} \leq \mu \leq \bar{x} + z\ \sigma_{\bar{x}} \} = 1 - \alpha$$

Proporção de Ocorrência (p)

$P_r \{ \text{‘}p - z \;\; \sigma_p \le p \le \text{‘}p + z \;\; \sigma_p \} = 1 - \alpha$

Para população finita usar se preciso o fator de correção: N – n/(N – 1) para o erro padrão da média e da proporção de ocorrência.

Para pequenas amostras ($n \le 30$ unidades ou $n \le 45$) substituímos os valores de "z" pelo o de "t" da distribuição de Student que se assemelha à distribuição normal padrão N (0 ; 1). Também usamos a distribuição "t", quando a variância populacional ($\sigma^2$) é desconhecida.

A variável "t" possui (n – 1) graus de liberdade.

$$P_r \left\{ \bar{x} - t \frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t \frac{s}{\sqrt{n}} \right\} = 1 - \alpha$$

Quando se trata de proporção, não se trabalha com amostra (n) pequena, daí não haver necessidade de usar a distribuição "t" de Student, mas somente a distribuição normal.

**Estimativa do total da característica observada** em uma população, através de dados amostrais, pode usar os processos de estimativa por ponto ou por intervalo.

Estimativas por ponto:

Variável continua: $\hat{E}t = Nx$ - Variável discreta: $\hat{E}t = N'p$

**Estimativas por Intervalo** – utilizamos os limites do intervalo de confiança calculado para estimativa do parâmetro médio da população e multiplicamos por N.

Exercícios Resolvidos

1. Com a finalidade de conhecer, com precisão de 95%, a idade média de 9.000 alunos da UFS, estudamos uma amostra de 400 alunos, registrando média de 23 anos e desvio padrão de 2 anos. Qual o intervalo de confiança para a média desta população (μ).

$$P_r \{ \bar{x} - z\,\sigma_x \le \mu \le \bar{x} + z\,\sigma_x \} = 1 - \alpha$$

$$P_r \left\{ 23 - 1{,}96 \frac{2}{\sqrt{400}} \le \mu \le 23 + 1{,}96 \frac{2}{\sqrt{400}} \right\} = 95\%$$

$$P_r \{ 22{,}8 \le \mu \le 23{,}2 \} = 95\%$$

Existe: 95% de probabilidade que a média da população esteja neste intervalo.

2. Com a finalidade de conhecer a percentagem de pessoas vacinadas contra varíola, em uma população de 70.000 habitantes, estudamos uma amostra de 1200 pessoas, das quais 960 estavam vacinadas. Qual o intervalo de confiança para esta proporção na população com acerto de 95%.

$$P_r \{ \text{‘}p - Z\ \sigma_p \leq p \leq \text{‘}p + z\ \sigma_p \} = 1 - \alpha$$

$$P_r \left\{ 0,8 - 1,96\sqrt{\frac{0,8 \cdot 0,2}{400}} \leq p \leq 1,96\sqrt{\frac{0,8 \cdot 0,2}{400}} \right\} = 95\%$$

$$P_r \{ 0,76 \leq p \leq P\ 0,84 \} = 95\%$$

A percentagem de vacinados na população está entre 76% a 84% com 95% de certeza, ou seja, devemos ter entre 53.200 a 58.800 pessoas vacinadas.

3. A amostra a seguir: $x_i = \{9, 8, 12, 7, 9, 6, 11, 6, 10, 9\}$ representa as idades de 10 crianças selecionadas em determinada população. Construir um intervalo de confiança para média populacional ao nível de 95%.

Média amostral: $\bar{x} = \frac{\sum x_i}{n} \Rightarrow \bar{x} = \frac{87}{10} = 8,7$

Desvio Padrão amostral: $s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} \Rightarrow s = 2$

Como não sabemos o tamanho da população (N) não podemos usar o fator de correção para a variância amostral, neste caso dividimos por (n – 1). Podemos, também, usar o fator de correção para população infinita, onde:

$$\sigma^2 = \frac{n}{n-1} s^2$$

$1 - \alpha = 95\%$ - números de graus de liberdade= $9 \Rightarrow t_{0,025} = \pm\ 2,2622$

$$P_r \left\{ \bar{x} - t \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t \frac{s}{\sqrt{n}} \right\} = 1 - \alpha$$

$$P_r \left\{ 8,7 - 2,2622 \frac{2}{\sqrt{10}} \leq \mu \leq 8,7 + 2,2622 \frac{2}{\sqrt{10}} \right\} = 95\%$$

$$P_r \{ 7,27 \leq \mu \leq 10,13 \} = 95\%$$

Existe 95% de probabilidade de μ está contida neste intervalo.

## ATIVIDADES

1. Uma amostra de 400 quilos de carne irradiada apresenta um índice médio de desidratação de 172 com desvio padrão 7,3. Obter o intervalo de confiança para a média populacional com 95% de probabilidade.

2. Uma amostra aleatória 400 estudantes de uma Universidade apresentou 96 deles devendo disciplinas de seus currículos. Adotando-se o nível de confiança de 95%, pergunta-se: qual a precisão de estimativa para a proporção de universitários que devem disciplinas em seus currículos?

3. Um conjunto de 15 animais apresentou os seguintes pesos: 30 , 22 , 32 , 26 , 30 , 24 , 40 , 34 , 36 , 32 , 33 , 28 , 30 , 32 , 32. Estimar a média dessa população com precisão de 95%.

4. Uma amostra de 100 crianças de 6 a 12 anos, do grupo escolar Nilo Peçanha – SC apresentou quanto ao índice CPOD um valor médio igual a 9 e variância de 6,25. Determinar o intervalo de confiança com 99% de probabilidade.

5. Em 100 suínos da raça Duroc, constatamos um peso médio da glândula adrenal igual a 3,68g. Apresentar o intervalo de confiança ao nível de 1% de probabilidade, sabendo-se que o coeficiente de variação encontrado foi igual a 18%.

6. Foi testada uma amostra de 1.600 cigarros de certa marca, com relação ao conteúdo de nicotina, sendo encontrado: $\overline{x} = 22$ e $s^2 = 16$. Calcule os limites de confiança de nicotina para essa população de cigarros com 99% de probabilidade

7. Em 900 cordeiros da Fazenda Pica-Pau Amarelo, Município de Santa Cruz – RJ foram verificados os seguintes dados hematológicos: pH médio do sangue: 7,40 e variância: 0,49. Obter o intervalo de confiança ao nível de 95% de probabilidade

8 – Suponha que as alturas dos alunos de nossa faculdade tenham distribuição normal com s = 15 cm. Foi retirada uma amostra aleatória de 100 alunos obtendo-se média igual a 175 cm. Construir, ao nível de confiança de 95% o intervalo para a verdadeira altura média dos alunos.

## CONCLUSÃO

A partir desta aula você deve saber o que é população e amostra. Deve conhecer a relação entre população e amostra, alem disto definindo o uso da amostragem na investigação de determinada população, você já aprendeu a fazer um planejamento sobre a realização de todas as etapas da pesquisa, bem como utilizar alguns elementos básicos nesta investigação, como: cadastro da população, aplicação de uma pesquisa piloto, elaboração do questionário, coleta dos dados, crítica e análise.

Você saberá que os elementos de uma população podem ser formados por: indivíduos, firmas, propriedades rurais, escolas, produtos, ou qualquer coisa que possa ser mensurada, contada ou ordenada segundo postos.

Saberá também que os termos populações e amostras se referem a um conjunto específico de circunstâncias. Ou seja, em determinado caso os alunos de uma sala de aula podem ser considerados como uma população, da qual iremos extrair amostras. Em outra situação, os mesmos alunos podem ser considerados como uma amostra de todos os alunos daquela universidade.

Com esta aula, você é capaz de entender que as características observadas por uma amostra devem ser utilizadas para estimar as mesmas características na população desta amostra, de acordo com determinado nível de confiança, são as estimativas de parâmetros médios populacionais por intervalo.

No final da aula temos uma lista de exercícios para serem resolvidos em grupos de no máximo cinco pessoas ou individual. Com certeza você vai ficar muito satisfeito com os resultados do seu desempenho.

## RESUMO

Nesta aula tratamos dos conceitos básicos da amostragem, além de explorar as razões que justificam a aplicação deste método de pesquisa, bem como os métodos de amostragem mais utilizados. Embora tenhamos alguns métodos de pesquisas por amostragem demos especial atenção à amostragem aleatória simples em decorrência de sua importância na análise estatística. Embora nenhum método possa garantir que a amostra seja exatamente semelhante à população da qual foi extraída uma amostra aleatória, a amostragem aleatória simples além da importância acima mencionada é de mais fácil aplicação e permite estimar o valor do erro amostral, isto é: dizer "quão próxima" está à amostra da população. As amostras não aleatórias não apresentam esta característica.

A amostra piloto foi outro tema tratado, mostrando o quanto ela é útil para avaliar o comportamento das principais características da população a ser investigada, para que você tenha uma base de dados que lhe permita calcular alguns parâmetros com a média aritmética, proporção de ocorrência, desvio padrão que serão utilizados para dimensionar o tamanho da amostra a ser pesquisada. A amostra piloto também é muito importante para testar o instrumento de coleta a ser utilizado na pesquisa. É uma grande oportunidade que você tem de modificar a forma de fazer alguma investigação ou mesmo incluir outras perguntas, bem com retirar aquelas que julgou serem inconsistentes.

Encerrando esta aula você tomou conhecimento de como estimar os parâmetros médios de uma população por ponto e por intervalo, sendo este último calculado não apenas com base nos parâmetros médios da amostra, mas com maior segurança pela utilização do erro de estimativa. Erro este que basicamente depende de três fatores fundamentais em qualquer pesquisa por amostragem: nível de confiança desejada, dispersão na população e tamanho da amostra.

## AUTO-AVALIAÇÃO

Sou capaz de trabalhar com variáveis aleatórias contínuas?

Sou capaz de calcular os parâmetros de uma variável aleatória contínua e entender seu significado?

Sou capaz de aplicar o modelo da distribuição de probabilidades Normal a partir do uso da normal padronizada?

Sou capaz de aplicar o modelo da distribuição de probabilidades Normal como aproximação da Binomial?

## PRÓXIMA AULA

Amostragem: Dimensionamento de amostras. Seleção dos elementos de uma amostra. Estimativa da característica total da população investigada.

## REFERÊNCIAS

RODRIGUES, PEDRO CARVALHO. **Bioestatística**. Universidade Federal Fluminense.
FONSECA, JAIRO DA. **Curso de Estatística**. Editora Atlas.
OLIVEIRA, FRANCISCO ESTEVAM MARTINS DE. **Estatística e Probabilidade**. Editora Atlas.
TANAKA. **Elementos de Estatística**. Editora McGraw.Hill.
BARBETTA, PEDRO A. **Estatística Aplicada as Ciências Sociais**. Editora da UFSC.
GÓES, LUIZ A. C. **Estatística I e II**. Editora Saraiva.
DÍAZ, FRANCISCA; LOPES, FRANCISCO JAVIER. **Bioestatística**. Editora Thomson.